Deep Transfer Learning for Crop Yield Prediction with Remote Sensing Data

Anna X. Wang
Stanford University
Department of Computer Science
annaxw@cs.stanford.edu

Caelin Tran
Stanford University
Department of Computer Science
caelin@cs.stanford.edu

Nikhil Desai
Stanford University
Department of Computer Science
nikhild@cs.stanford.edu

David Lobell
Stanford University
Department of Earth System Science
dlobell@stanford.edu

Stefano Ermon
Stanford University
Department of Computer Science
ermon@cs.stanford.edu

ABSTRACT

Accurate prediction of crop yields in developing countries in advance of harvest time is central to preventing famine, improving food security, and sustainable development of agriculture. Existing techniques are expensive and difficult to scale as they require locally collected survey data. Approaches utilizing remotely sensed data, such as satellite imagery, potentially provide a cheap, equally effective alternative. Our work shows promising results in predicting soybean crop yields in Argentina using deep learning techniques. We also achieve satisfactory results with a transfer learning approach to predict Brazil soybean harvests with a smaller amount of data. The motivation for transfer learning is that the success of deep learning models is largely dependent on abundant ground truth training data. Successful crop yield prediction with deep learning in regions with little training data relies on the ability to fine-tune pre-trained models.

CCS CONCEPTS

• **Applied computing** → *Environmental sciences*;

KEYWORDS

Sustainability, agriculture, deep learning

ACM Reference Format:

Anna X. Wang, Caelin Tran, Nikhil Desai, David Lobell, and Stefano Ermon. 2018. Deep Transfer Learning for Crop Yield Prediction with Remote Sensing Data. In *COMPASS '18: ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS), June 20–22, 2018, Menlo Park and San Jose, CA, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3209811.3212707

COMPASS '18, June 20–22, 2018, Menlo Park and San Jose, CA, USA

ACM ISBN 978-1-4503-5816-3/18/06…$15.00
https://doi.org/10.1145/3209811.3212707

1 INTRODUCTION

Accurate forecasts of crop production facilitate essential tasks like determining the optimal profile of crops to plant, allocating government resources, effective planning and preparation of aid distribution, and decision-making about imports and exports in more commercialized systems. Currently, leading crop yield prediction techniques mostly rely on locally sensed data, such as rainfall measurements and farmer surveys from field visits. Locally sensed data provide detailed information but are expensive to collect, often noisy, and extremely difficult to scale. Remote sensing data, a cheap and globally accessible resource, coupled with modern machine learning approaches offers a potential solution [3, 6, 12, 13]. Our work in this paper focuses on the task of predicting Argentine and Brazilian soybean harvests with deep learning models trained on remote sensing data.

1.1 Related Work

Prior models for predicting crop yields from remotely-sensed data utilize human-engineered features, such as NDVI [7], EVI2 [2], and GCVI [4], which represent specific combinations of measurements on one or two spectral bands. While such features are useful for visualization and on-the-ground analysis, they discard large amounts of spectral information from particular images and may not represent the optimal feature combination for machine learning.

In recent work, You et al. [13] applied convolutional and recurrent neural networks to the yield estimation task, allowing more flexible use of information in all spectral bands. The authors made the key simplifying assumptions that the pixels of input images may be considered permutation-invariant and spectral bands may be considered uncorrelated. This enabled the usage of a histogram of per-band pixel counts as input in lieu of the raw satellite image, generating significant dimensionality reduction while preserving sufficient information. Working with data for 8,945 harvests in total, the authors tested their models on corn and soybean yields in the United States from the years 2011 to 2015 using a model trained on data from all preceding years, and achieved an average 5.55 bushels/acre root mean square error (RMSE).

Here we extend the approach in You et al. to new regions, specifically Argentina and Brazil. Unlike their study, which tested the model in the same region as it was trained, we test the ability to transfer a model trained in one region to another. The motivation for transfer learning is that the success of deep learning models is largely dependent on the abundance of ground truth training data. The ability to achieve successful crop yield predictions in developing countries with fewer available data points requires the ability to fine-tune pre-trained models from countries where data is more readily available.

1.2 Dataset and Features

To perform the crop yield prediction task with remotely sensed data, we leveraged Moderate Resolution Imaging Spectroradiometer (MODIS) satellite imagery, which provides free and easy to access coverage of the entire globe. Of the many imagery products exported by MODIS, we used MOD09A1 [10], MYD11A2 [11], and MCD12Q1 [1], which provide, respectively, eight-day composites for seven-band reflectance imagery, two-band daytime and nighttime temperature imagery, and a land cover mask. Each reflectance band represents a distinct range of wavelengths sensed by the MODIS satellite. The land cover mask is updated annually and was only used to distinguish cropland from non-cropland.

As our ground truth for soybean crop yields, we used county-level and province-level yield statistics compiled by the Argentine Undersecretary of Agriculture [9] and the Brazilian Institute of Geography and Statistics [8]. All yields were reported in units of metric tonnes per cultivated hectare (t/Ha).

Each officially reported crop yield of a particular region for a particular harvest was paired with a sequence of MODIS reflectance and temperature images from the months preceding the harvest. In order to train our deep learning models, we processed the MODIS imagery into the dimensionally reduced pixel histograms described by You et al. For each image $I \in \mathbb{R}^{h \times w \times d}$, where h, w, and d are respectively the number of image height pixels, width pixels, and reflectance/temperature bands, we assumed each band to be independent of all others and created d pixel histograms. For each histogram, we generated a number of buckets that represented different ranges of reflectance values or temperatures and placed pixels into their corresponding buckets. We grouped these individual histograms together, so the final representation of I was a two-dimensional matrix $H \in \mathbb{R}^{b \times d}$, where b is the number of bins we chose. For each soybean harvest, we stacked histogram matrices from the image sequence associated with that harvest into a single three-dimensional tensor.

Because the MODIS cropland mask does not distinguish soybeans from other crops, we ignored regions that contributed the bottom 5% of total production in Argentina and Brazil in order to only train our models on regions with significant soybean crop cover and also filter out noisy crop yield values from regions with very low soybean production. Unfortunately, this eliminated a large number of harvests from our datasets, increasing the difficulty of the yield prediction task, especially in Brazil. Our Argentine dataset contained 1,837 harvests after filtering, and our Brazilian dataset contained 336 harvests after filtering.

The relatively large volume of our data from Argentina and the country's geographical proximity to Brazil inspired the use of transfer learning to improve predictive performance. A preprocessing step necessitated by our transfer learning approach was to match the lengths of harvest image sequences between Argentina and Brazil. The soybean growing season in Argentina is October to June, while the soybean growing season in Brazil is September to April. In order to keep the time periods consistent, we ignored the first month of the Argentina growing season so that the input tensor, regardless of origin country, covered a period of roughly eight months.

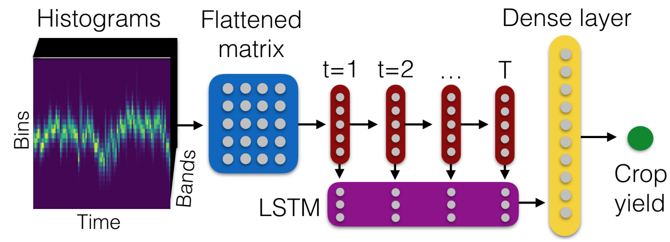

Figure 1: Visualization of the LSTM architecture

2 MODELS

2.1 Baseline Models

Our baseline models used ridge regression with varying regularization constants. Using histograms directly for ridge regression was infeasible due to their high dimensionality. We leveraged two different methods of feature extraction to reduce dimensionality. In the "band mode" method, we used only the mode of each band's histogram slice at each time step, creating an input vector of length dt for a single harvest. As an example, the first band of the histogram time series shown in Figure 1 would be replaced by a vector containing the bin index that is the brightest at each time step. Our other feature extraction method utilized the Normalized Difference Vegetation Index (NDVI) due to the metric's prevalence in industry. For each input sequence, we calculated the mean NDVI at each time step, yielding a feature vector of length t.

2.2 Deep Learning Models

Our primary model was a recurrent neural network composed of long short-term memory (LSTM) cells. We flattened the three-dimensional pixel histogram representation of a region into a two-dimensional matrix by concatenating on the MODIS bands and histogram bucket dimensions, which preserved the time dimension. The LSTM layer took d histograms as input at each time step and sent its ultimate activation to a final dense layer, which output the predicted crop yield. A visualization of our model architecture is shown in Figure 1.

For transfer learning from Argentina to Brazil, we initialized the LSTM model with the parameters from a neural network trained on Argentine soybean harvests. We stripped out the last dense layer of the pre-trained model and replaced it with an untrained dense layer of the same dimensions before training the modified model on the available Brazilian training data. In this manner, we fully recalibrated the last dense layer and fine-tuned the rest of the Argentine model's parameters.

3 RESULTS AND DISCUSSION

3.1 Argentina

We trained and tested our models on Argentine soybean harvests from 2012 to 2016 in order to evaluate the efficacy of the pixel histogram and LSTM model approach in a developing country with less available data. For each testing year, we trained the model on harvests from all years except for that year. Learning rates and stopping criteria were tuned on a hold-out validation set sampled from 20% of the training data. A comparison of deep learning and baseline RMSE values can be found in Table 1 and the associated R^2 values for the LSTM models are in Table 3. On average, the LSTM models outperformed ridge regression baselines, demonstrating the utility of the approach.

Notably, we observed that 2014 is an outlier year with a negative R^2 score for both the neural network model and the baselines. Negative R^2 scores occur when a model does not follow the trend of the data. This performance was likely due to the fact that the test set was not sampled randomly from the full population of harvests from all years but was instead sampled from a single year, in this case 2014. Any anomalies localized to that year, such as unusual weather patterns like the onset of the strong 2014-2016 El Niño event [5] or social factors, could have disproportionately impacted soybean yields and led to poor generalization to this specific test year.

In addition, we trained our model to forecast Argentine soybean crop yields in advance of the harvest date. For example, in order to forecast the soybean yield four months (about 50% of the season) in advance of the harvest in June, we withheld the second half of the image sequence corresponding to that harvest, training and predicting only using the first half. Figure 2 shows the results of this forecasting strategy with season fractions ranging from 25% to the full 100%. Performance on the test year was in general best with the full data available during training, but even with access to only 25% of the information, our model exhibited positive predictive power in many years.

As expected, we saw that the predictive performance of our model on the training and development sets increased monotonically as we provided more data. However, this was not true of the test set in all cases. Namely, there were sometimes dips in R^2 (or, equivalently, jumps in RMSE) when predicting with 50% and 75% of the data. Once again, this was likely due to the nonrandom nature of the test set. It is possible that providing additional mid-season data helped the model learn features that were important to the training and development sets but were irrelevant to the given test set. Interestingly, performance sometimes improved when 100% of the season data was provided even after experiencing dips with 50% and/or 75%. This suggests that the beginning and end conditions of a harvest are the most consistent indicators of yield between years.

3.2 Brazil

Given the promising performance in Argentina, we turned our attention to Brazilian soybean harvests to test our model's transferability to different regions. As reference points, ridge regression baselines and an LSTM model were trained on only Brazilian soybean harvests from 2012 to 2016 while a separate LSTM model used transfer learning from Argentina.

Our transfer learning model outperformed all other models that were trained only on Brazilian data. A comparison of RMSE values are shown in Table 2. A more detailed breakdown of RMSE and R^2 scores for our standard and transfer learning models is shown in Table 3.

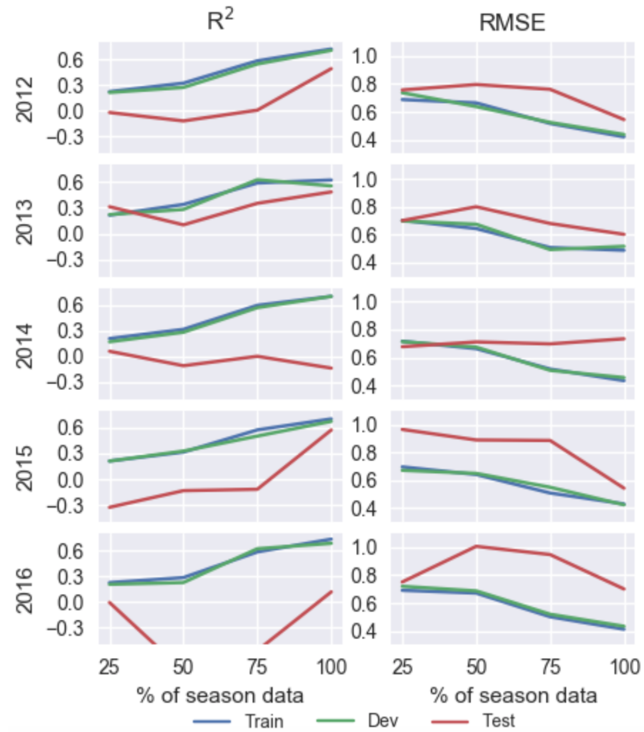

Figure 2: Argentine soybean harvest forecast performance on test years 2012-2016 as a function of season data

Harvest	LSTM; histograms	Regression; NDVI	Regression; band modes
2012	0.54	0.60	0.64
2013	0.60	0.59	0.67
2014	0.73	0.75	0.75
2015	0.54	0.94	0.93
2016	0.70	0.92	1.04

Table 1: RMSE of LSTM and baseline models for Argentine experiments

Harvest	LSTM; histograms	LSTM; histograms & transfer learn	Regression; NDVI	Regression; band modes
2012	0.42	0.38	0.56	0.68
2013	0.29	0.26	0.40	0.60
2014	0.23	0.26	0.28	0.33
2015	0.53	0.50	0.54	0.60
2016	0.62	0.52	0.49	0.73

Table 2: RMSE of LSTM and baseline models for Brazilian experiments

Model training	Test country	Harvest	R^2	RMSE
Argentina	Argentina	2012	0.49	0.54
		2013	0.49	0.60
		2014	-0.13	0.73
		2015	0.57	0.54
		2016	0.12	0.70
Brazil	Brazil	2012	0.57	0.42
		2013	0.50	0.29
		2014	0.46	0.23
		2015	-2.03	0.53
		2016	0.043	0.62
Brazil + transfer learning from Argentina	Brazil	2012	0.66	0.38
		2013	0.58	0.26
		2014	0.36	0.26
		2015	-1.75	0.50
		2016	0.33	0.52

Table 3: R^2 and RMSE of deep learning model predictions on Brazilian and Argentine soybean harvests

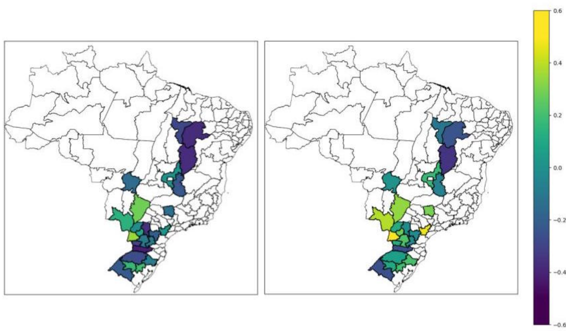

Figure 3: Error of native model trained on Brazilian data (left), and of transfer-learning model trained on Argentine and Brazilian data (right). Error is defined as (predicted yield) − (true yield).

Analysis (Figures 3 and 4) of our transfer learning model's performance on test data from a different geographic location than its original training data was promising. Most significantly, using the parameters of our Argentine model as initialization for training led to four-fold speedup in convergence time and improved performance on the Brazil test set.

Geographic visualization of model error indicated that our transfer learning model outperformed our initial model in the most fertile soybean-producing regions of Brazil, which are the states of Rio Grande do Sul, Santa Catarina, and Paraná. This reflects our intuition that model insights from learning in Argentina persist in the transfer learning model and helps improve performance in regions in Brazil similar to those in the Argentine training set. Interestingly, we filtered out imagery from the most geographically close provinces of Argentina to Brazil, the "Mesopotamian" provinces of Corrientes

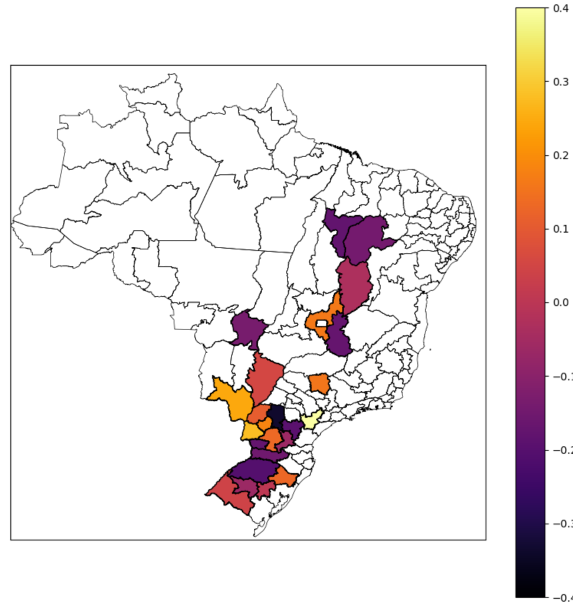

Figure 4: Improvement of transfer-learning model over native model. Improvement is defined as |transfer model error| − |native movel error|. Darker colors indicate improved performance of the transfer-learning model compared to the native model.

and Misiones, during training because of low crop production. Future work may consider including training data from these provinces for a possible performance improvement in Brazil.

4 CONCLUSIONS AND FUTURE WORK

This paper presents a preliminary deep transfer learning framework for reliable crop yield prediction in developing countries with remote sensing data. The results in Argentina and Brazil demonstrate that this approach can successfully learn effective features from raw data and achieve improved performance compared to traditional methods. The ability to improve predictive performance in regions with limited data by using transfer learning is exciting because these regions especially stand to benefit from a cheap, reliable crop prediction tool. Next steps include expanding the application of this approach to new regions, supporting more crops, and using models pre-trained on the United States, which has a significant amount of reliable data, to transfer learn to other countries.

REFERENCES

[1] 2013. MCD12Q1 MODIS/Land Cover Type Yearly L3 Global 500m SIN Grid. (2013).

[2] Douglas K. Bolton and Mark A. Friedl. 2013. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology* 173 (2013), 74–84. https://doi.org/10.1016/j.agrformet.2013.01.007

[3] Neal Jean, Marshall Burke, Michael Xie, W Matthew Davis, David B Lobell, and Stefano Ermon. 2016. Combining satellite imagery and machine learning to predict poverty. *Science* 353, 6301 (2016), 790–794.

[4] David B. Lobell, David Thau, Christopher Seifert, Eric Engle, and Bertis Little. 2015. A scalable satellite-based crop yield mapper. *Remote Sensing of Environment* 164 (7 2015), 324–333. https://doi.org/10.1016/j.rse.2015.04.021

[5] Michael McPhaden and Aaron Levine. 2016. NOAA: How the failed 2014-15 El Nino fueled the strong 2015-16 El Nino. https://www.pmel.noaa.gov/elnino/news-story/how-failed-2014-15-el-nino-fueled-strong-2015-16-el-nino

[6] Reid Pryzant, Stefano Ermon, and David Lobell. 2017. Monitoring Ethiopian Wheat Fungus with Satellite Imagery and Deep Feature Learning. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 1524–1532.

[7] N. A. Quarmby, M. Milnes, T. L. Hindle, and N. Silleos. 1993. The use of multi-temporal NDVI measurements from AVHRR data for crop yield estimation and prediction. *International Journal of Remote Sensing* 14, 2 (1993), 199–210. https://doi.org/10.1080/01431169308904332

[8] Brasil Sistema IBGE de Recuperacao Automatica, Instituto Brasileiro de Geografia e Estatistica. [n. d.]. Producao Agricola Municipal: producao das lavouras temporarias. https://sidra.ibge.gov.br/tabela/1612

[9] Delegaciones y Estudios Economicos Argentina Subsecretaria de Agricultura, Direccion Nacional de Estimaciones. [n. d.]. Datos Argentina: Estimaciones agricolas. http://datos.gob.ar/dataset/estimaciones-agricolas

[10] Eric Vermote. 2015. MOD09A1 MODIS/Terra Surface Reflectance 8-Day L3 Global 500m SIN Grid V006. (2015). https://doi.org/10.5067/modis/mod09a1.006

[11] Zhengming Wan and S. Hook. 2015. MYD11A2 MODIS/Aqua Land Surface Temperature/Emissivity 8-Day L3 Global 1km SIN Grid V006. (2015). https://doi.org/10.5067/modis/myd11a2.006

[12] Michael Xie, Neal Jean, Marshall Burke, David Lobell, and Stefano Ermon. 2016. Transfer learning from deep features for remote sensing and poverty mapping. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*. AAAI Press, 3929–3935.

[13] Jiaxuan You, Xiaocheng Li, Melvin Low, David Lobell, and Stefano Ermon. 2017. Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data. *2017 Association for the Advancement of Articial Intelligence* (2017).